Report by: Anand Prashar

NLP Homework 5 – q7

---

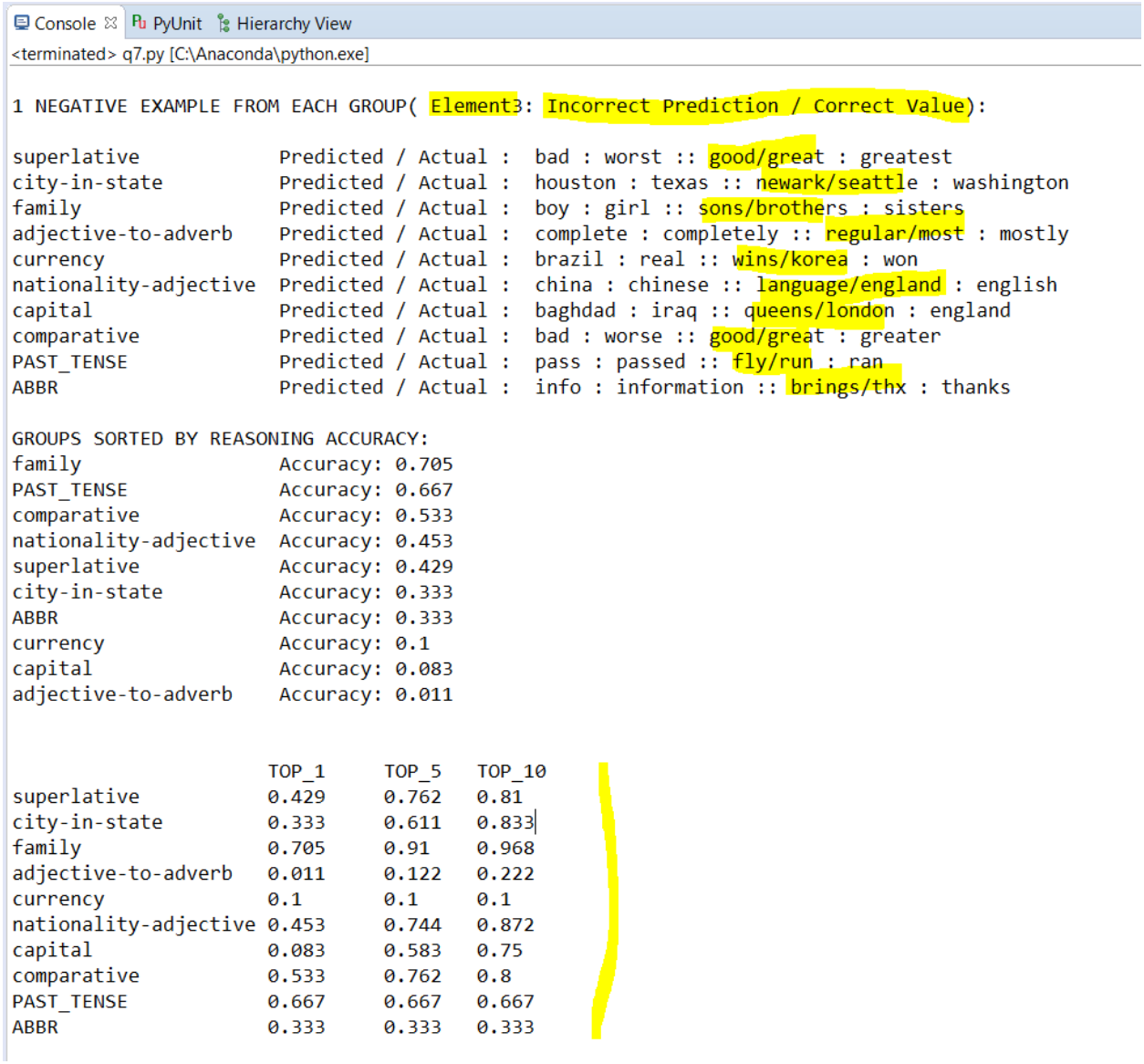
Q:

For each of the eight relation groups, print the 1-best, 5-best, and 10-best accuracy of your vectors on the group. The n-best accuracy is the percentage of items for which the correct answer was in the top n vectors returned. Use the approach from the sample code in Question 6 to complete the task. That is, given $w_1$, $w_2$, $w_3$, $w_4$, calculate $w_1 - w_2 + w_4$, obtain the closest words, and compare them to $w_3$.
On CrowdMark, in the space below, display the table generated by your q7.py program. Additionally, for each relation group, show an example of an incorrectly predicted analogy item, along with the correct answer.
Are there certain kinds of relations that seem to be predicted more accurately or less accurately by this method? Discuss your opinions and give a rationale for them.

---

A:

Below is the snapshot Q7.py ( next page)




Explanation-

In First block, I am fulfilling the requirement of printing 1 wrong classification of each Relation/Group

In Second block, I am printing Groups in sorted order of their precision ( == TOP_1 )
It helps me see which Relations are predicted more accurately

In Third block, I am printing TOP_1, TOP_5, TOP_10 matrix as asked.

*Please ignore relations*: PAST_TENSE, ABBR ( they were for q8 analysis appended by me )

Yes, certain relations seem be more accurately predicted than others. Looking at second chart of reasoning accuracy, we see 'family' has considerably high accuracy. My understanding would be there

are not as much vectors in this domain AND the vectors is a common noun too. It tends to be predicted more accurately. Similarly, ‘Capital’ category contains words which do not typically occur in same context causing defined observation of V1-V2+V4 to not be satisfactory.